Water Resources Management **10:** 219–240, 1996.

Application of First-Order and Monte Carlo Analysis in Watershed Water Quality Models

A. GHOSH BOBBA[1] , VIJAY P. SINGH[2] and LARS BENGTSSON[3]
[1] *National Water Research Institute, Burlington, Ontario, Canada, L7RT 4A6.*
[2] *Department of Civil and Environmental Engineering, Lousiana State University, Baton Rouge, LA 70803–6405, U.S.A.*
[3] *Department of Water Resource Engineering, University of Lund, Box 118, S-22100 Lund, Sweden.*

(Received: 29 November 1994; in final form: 26 October 1995)

Abstract. To achieve effective environmental control, it is important to develop methodologies for dealing with uncertainties in model simulation of pollution behaviour and effects. Several procedures have been proposed to quantify uncertainties in modelling studies. This paper utilizes the two methods that are widely applied, i.e. functional analysis and Monte Carlo Simulation.

The first-order part of the functional analysis method provides a measure of uncertainties in dependent variables in terms of uncertainties in independent variables. The procedure is based on first-order terms in the Taylor series expansion of the dependent variable about its mean value with respect to one or more independent variables. The major assumption in this procedure is that all independent and dependent variables are the second moment variables (SMV), which means that the behaviour of any SMV is completely described by its mean and standard deviation. The mathematical simplicity of the procedure allows application by simple input-output models. Consequently, it has been applied to many environmental simulators, e.g. hydrological models, stream water quality models, lake water quality models and ground water pollution models.

The Monte Carlo Simulation (MCS) method uses a large number of repeated trials or simulations with the values for stochastic inputs or uncertain variables selected at random from their assumed parent probability distributions to establish an expected range of model uncertainty.

Key words: Water quality modelling, uncertainty analysis, function analysis, Monte Carlo method.

1. Introduction

Numerous water quality models have been developed to simulate of physical and biological processes occurring in stream waters. Applications range from identifying the processes affecting stream water quality to forecasting the quality for operational purposes.

It has been a common practice to describe the problems related to chemical and biological processes in river waters through deterministic partial differential equations (**PDE**). These models are deterministic in that they provide a single response for each set of model parameters and initial conditions. However, there is always uncertainty, both in the evaluation of field data and in the use of mathematical models to predict the outcome of natural processes, since the processes are still poorly understood and full mathematical representation is usually too complicated and too costly to implement. In general, one must deal with four aspects of

uncertainty when modelling water quality (Beck, 1987): (a) uncertainties about model structure, i.e., what basic processes are involved, how they interact, and how these processes and interactions can be mathematically characterized in an efficient and parsimonious manner; (b) uncertainty in the model parameters, i.e., parameter identification and calibration problems; (c) uncertainty associated with prediction of the future behaviour of the system, i.e. aggregation of model structure and model parameter uncertainties into overall prediction uncertainty; and (d) reduction of critical model uncertainties via carefully designed experiments and/or monitoring programs. There is also inherent variability and randomness in natural processes and their measurements. All these sources of uncertainty may be represented as input forcing terms in the balance equations. The initial conditions may also be random, either because knowledge of the real initial conditions is imperfect or because measurements are affected by random variations. The model coefficients (rate constants) are random either because our assessment is not perfect or because of random variations in measurements. Inputs may also be uncertain because estimates of future loadings, based on projections and future loadings, may be biased.

The above four types are strongly intertwined. For example, type (c) assumes that the model structure (type (a)) has been reasonably identified, and examines the propagation of model parameters (type (b)) and data uncertainties through the model and the resulting prediction uncertainty. Further, the techniques employed in type (c) may also identify the critical sources of model uncertainty that type (d) seeks to reduce. While each type is of great importance and their interactions cannot be ignored, type (c) is of primary importance from the point of view of the decision maker. The decision-making processes often cannot wait for better models to be developed or new data to be obtained. Water quality standards must be set and management strategies implemented in light of existing models and data. Hence, assessment of prediction uncertainty has assumed a much larger role in water quality decision-making in recent years. Fortunately, there are sound, well tested mathematical methods available for prediction-uncertainty estimation. These methods are first-order uncertainty and Monte Carlo Simulation. This paper will describe the application of first-order error analysis (FOEA) and Monte Carlo Simulation (MCS) analysis to a water quality model.

2. Functional Analysis

The title 'functional analysis' refers to any method which involves the use of an assumed function to approximate the mean, variance and/or higher-order moments of an output variable (Y) as a function of one or more input variables (x). This grouping encompasses several closely related methods identified in the literature as error analysis, uncertainty analysis, or confidence interval development. A representative example of a functional analysis method is the method of Benjamin and Cornell (1970) which is based on using the first derivative terms in the Tay-

lor series expansion of the dependent variable. If it is assumed that a functional relationship exists between one dependent variable, Y, and multiple independent random variables ($x_1, x_2,...,x_n$) as:

$$Y = f(x) = f(x_1, x_2, ..., x_n). \tag{1}$$

Equation (1) can be linearized by writing the Taylor series expansion about the mean value of each independent variable; the mean and variance of Y can be obtained from the linearized equation as follows:

$$E(Y) = f(\mathbf{x}). \tag{2}$$

$$V(Y) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial f(x)}{\partial x_i} \frac{\partial f(x)}{\partial x_j} \operatorname{cov}(x_i x_j), \tag{3}$$

where $\mathbf{x}$ denotes the mean vector of the independent random variable and $E(Y)$ and $V(Y)$ denote expected value and variance, respectively. All derivatives are evaluated at $\mathbf{x}$. Equation (2) then implies that the expected value of Y, $E(Y)$, is the functional value of the mean values of all the independent variables.

If all input variables are mutually independent (i.e. uncorrelated), all covariances between any two different random input variables will disappear from Equation (3), and the variance of Y can be rewritten as

$$V(Y) = \sum_{i=1}^{n} \left[\frac{\partial f(x)}{\partial x_i} \right]^2 V(x_i). \tag{4}$$

Equation (4) can be used to estimate the effect of the uncertainty in each model parameter and input variable on the model output. The first-order derivatives, $\partial f(x)/\partial x_i$, are also called sensitivity coefficients because they describe the variation in the dependent variable caused by a small change in each independent variable.

This approach has the advantage of only requiring estimates of the mean and variance (first and second moments) of the input parameters. In situations where limited information restricts analysis of the random components, the first-order analysis is a useful tool. In some cases, the first order-analysis gives answers comparable to those obtained by more complex analysis procedures. However, in more complicated situations, there are discrepancies between first-order analysis and nonlinearized methods such as a Monte Carlo technique (Scavia *et al.*, 1981). Other functional analysis approaches, such as statistical estimation of moment generating equations (Karmeshu and Lara-Rosano, 1987), point estimation techniques (Thorburn, 1986), or numerical analysis (Dresnack and Dobbins, 1968) are also usually limited to first and second moments.

3. Monte Carlo Simulation

Monte Carlo Simulation methods were developed in the 1940's as a numerical tool to solve complex sets of equations which were beyond the computational power

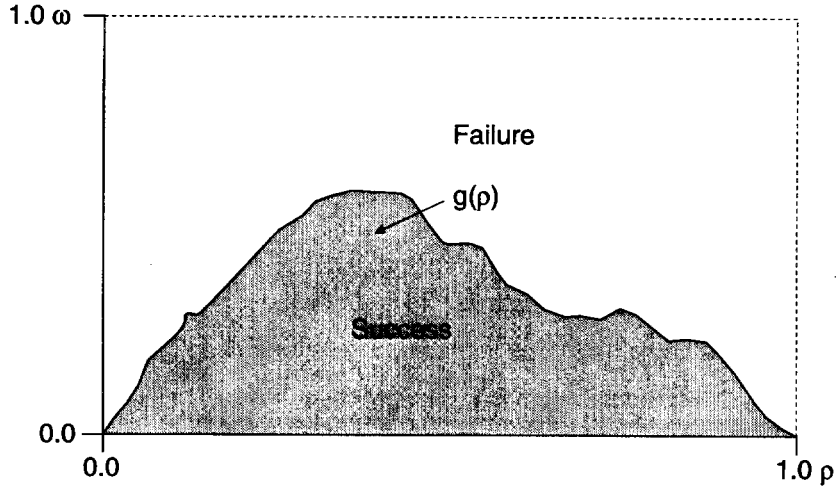

Figure 1. Application of Monte Carlo technique (from Khan 1950).

available at the time (Ulam and von Neumann, 1945; Metropolis and Ulam, 1949). Monte Carlo Simulation (MCS) is a conceptually simple process that has gained popularity with the development of high speed computers. For this method, a probability distribution type is identified or assumed for each uncertain characteristic in the model. This includes model equations, input variables, and parameters. The model is then run repeatedly for predetermined number of time, each time with the value for each uncertain characteristic being randomly selected from the probability distribution. The result is a distribution of values for the model prediction. This distribution can be regarded as the prediction error, representing the combined effects of all the uncertainties in characterizing the model.

Due to the characteristics of MCS, two important factors to be considered for its application are specifying the type of probability distribution of each random variable and selecting the total number of simulations to perform.

An accurate or appropriate description of the probability distribution of each random variable is necessary to properly utilize Monte Carlo Simulation (MCS), but for environmental engineering problems, certain distributions are more meaningful and fit the data better than others; these include Normal, Log Normal, Gumbel, Log-Gumbel, Gamma, Log-Pearson Type-III, Beta, Weibull, and Uniform.

Originally, the method was used as a numerical technique to solve differential and integral equations (Meyer, 1956). An excellent example of this application which elucidates the basic principles of Monte Carlo is given in a paper by Kahn (1950). Consider the problem of evaluating the following integral:

$$\int_0^1 g(x)\,\mathrm{d}x \; 0 < g(x) < 1. \tag{5}$$

The solution is to be determined using random sampling with the following rules: (a) ω and ρ are selected at random from a population uniformly distributed

between 0 and 1; and (b) if ρ is less than $g(\rho)$, the game is success, otherwise it is failure. Figure 1 shows that the probability of successes is equal to the area under the curve $g(\rho)$. Therefore an estimate for the solution of Equation (5) can be obtained by r/N, where r is the number of successes and N is the total number of trials. This example highlights the two basic traits of Monte Carlo methods: random sampling from a specified distribution and the use of that sample in a specified equation.

Monte Carlo methods are based on random sampling of a specified set. When used as a numerical technique in integration, the specified set is set of possible solutions to the equations over the range being examined. In environmental modelling, Monte Carlo Simulation methods are used to randomly select input values or equation parameters. Because the Monte Carlo Simulation method provides a straight forward way of including stochasticity in environmental models, it is probably the most widely used stochastic method. The procedure can easily be extended to several variables simultaneously. Input variables and/or parameters can be linked by predicting the selection of one variable upon the random outcome of another. Monte Carlo Simulation methods have been used to model environmental systems ranging in complexity from first order decay of organic chemicals in a river (Kothandaraman and Ewing, 1969; Burgers and Lettenmair, 1975) to highly diversified biological communities (Tiwari and Hobbie, 1976). The method has also been used in lake models (Fedra, 1979), predator-prey models (Gardner *et al.*, 1980), and ground water pollution models (Smith and Schwartz, 1981; Black and Freyburg, 1987). Results of Monte Carlo Simulation are a useful standard to which other methods are compared (van de Kramer, 1983; Malone *et al.*, 1983).

Because each input variable is randomly selected for each simulation, its value changes from simulation to simulation. However, its frequency distribution will resemble its parent distribution only when a sufficient number of simulations have been performed. In other words, the greater the number of simulations, better resemblance between generated and parent distribution of each random variable. This in turn results in more reliable information on model output frequency distributions, provided the appropriate input probability distribution is known.

Several methods are available to determine an adequate number of simulations. Malone *et al.* (1983) cited two methods. One method consists of iterating the sample generating procedure until the sample statistic converges to its population or historical value. The other approach directly calculates the required number of simulations based on the confidence interval for the mean of a process with known variance (Burgers and Lettenmair, 1975). Hahn and Shairo (1967) derived the following Equation (6) for estimating the number (n) of Monte Carlo samples that would be required to define the mean with desired error bounds.

$$n = \left[\frac{(Z_x \times S)}{y}\right]^2, \tag{6}$$

$$z_x = \frac{S}{\sqrt{n}} \leq y, \tag{7}$$

where z_x is a normal deviate at a confidence level, x; S is estimate of standard deviation of a selected variable; and y is the maximum desired error for a selected variable.

Another approach is to utilize the Kolmogorov-Smirnov (KS) goodness-of-fit statistic (Spear, 1970). Whitehead and Young (1979) used this sample in the Monte Carlo Simulation of a dynamic, stochastic model for water quality to obtain distributions for BOD and DO which closely corresponded with the true distribution.

4. Non-point Source Water Quality Model

SNSIM is a mathematical model which can be used to formulate a steady-state, one-dimensional, simulation model of a stream network (Braster and Chapra, 1973). SNSIM appears to be appropriate for application of FOEA and MSC error analysis methods. The reasons are that FOEA requires explicit expressions of mathematical models, i.e., analytical solutions, and SNSIM is capable of handling a variety of variables and parameters, such as non-point source waste loads, algal growth, and sediment oxygen demand. Using systems analysis, a mathematical model can easily be written to determine the levels of dissolved oxygen (DO), carbonaceous biochemical oxygen demand (CBOD) and nitrogen biochemical oxygen demand (NBOD) in a stream. It is designed to evaluate and or to predict the dissolved oxygen, and the carbonaceous and nitrogenous BOD profiles in a river or stream where the effects of dispersion can be assumed to be insignificant.

The stream network consists of a river and its tributaries Figure 2, which are segmented into sections of constant hydrologic, physical, chemical and biological parameters. Loads may be applied pointwise at the ends of the section or as distributed sources along its length. Also point sources of both BOD and DO deficit from minor tributaries can be input at the ends of a section and background loads of BOD and DO deficit can be introduced at the system's upstream ends. The mathematical equations for the calculations of the distribution of CBOD, NBOD and DO deficit have been explained by Braster and Chapra (1973).

An alalytical expression is then applied to each section to determine its CBOD, NBOD and DO deficit response to the loadings. This expression is similar in form to the classical Streeter-Phelps equation. Mass balance equations are applied at the junction of sections as well as the more complex junction of the system tributaries. In this simulation, the model generates results for the entire system.

A computer program was initially written for a single simulation of a river system (Braster and Chapra, 1973). For application of FOEA and MCS, the program was modified keeping most of features in the original program. The only modification made to the mathematical expressions in the carbonaceous BOD removal rate was that deoxygenation rate was assumed to be identical and that the (dis-

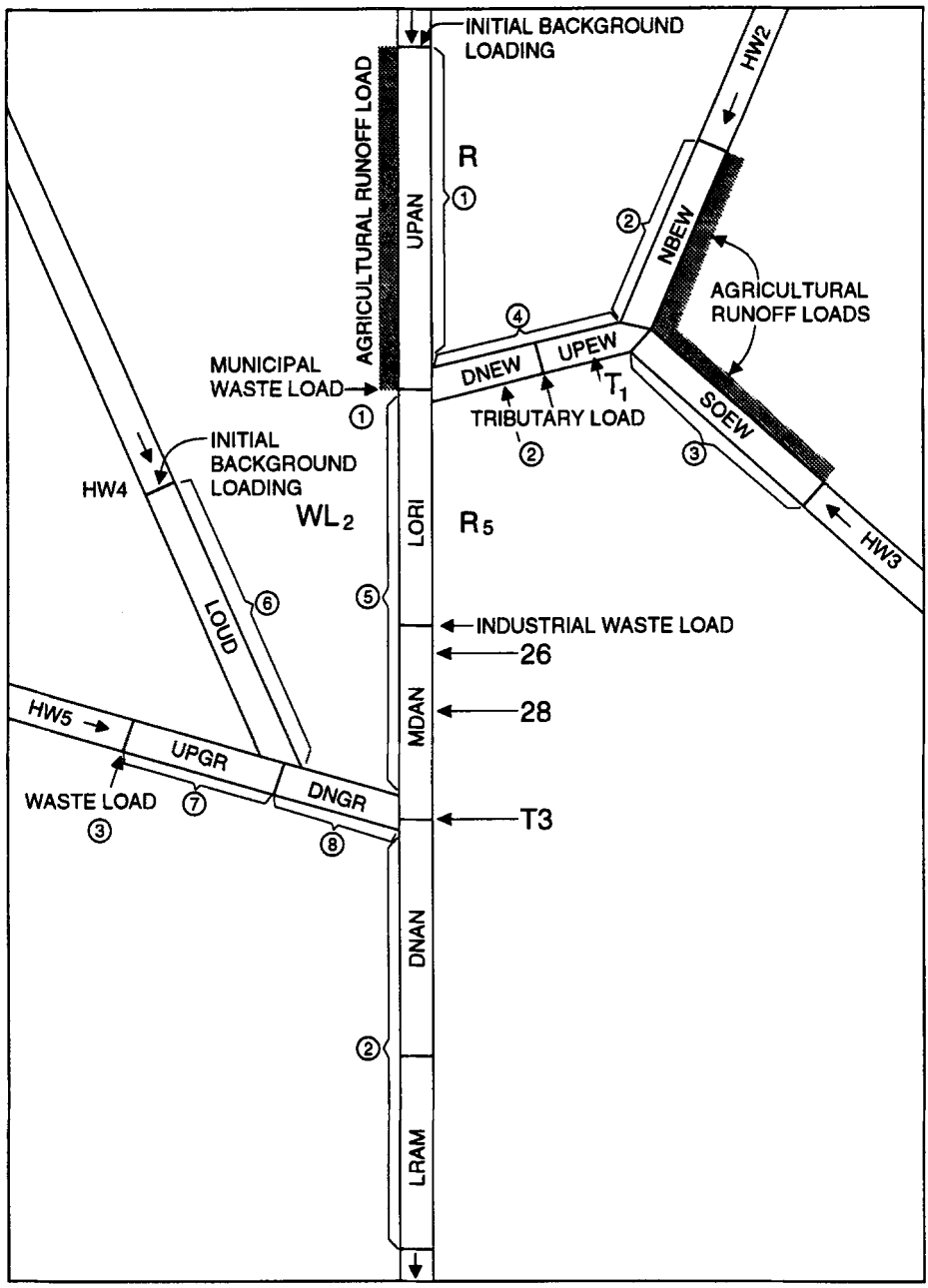

Figure 2. Schematics of river in watershed.

tance/velocity) term replaced with travel time. As an example, the restatement of the carbonaceous BOD equation:

$$L_{c,t} = L_{c,0}\, e^{(-k_c t)} + \frac{L_{c,d}}{k_c}(1 - e^{(-k_c t)}),$$

where

$$t = \frac{x}{u}. \tag{8}$$

Whenever the stream encounters different conditions, at any downstream point, say t_1, such as the presence of point loads, tributaries, or different values of reaction coefficients, and others, the model needs to replace the initial value terms by combining all the effects from upper reaches, tributaries and point loads. At that point, the uncertainty associated with each initial variable contains not only the uncertainty at the headwaters, but all the components contributed by uncertainty from tributaries and point loads. Then, further downstream, say t_2, uncertainty in any process variable contributes to total uncertainty in the dependent variable in two ways: (a) uncertainty of process variable from the head water t_1 is included in uncertainty terms of dependent variable at t_1, which then serves as the independent variable from t_1 to t_2, directly contributes to the uncertainty in dependent variables; (b) the sum of these two components is the total contribution to the dependent variable at t_2 by a process variable.

In the modified model, FOEA is available as an option, and up to 16 variables can be randomized which includes: three reaction coefficients, (k_c, k_n and k_a); algal oxygen production rate; sediment oxygen demand, headwater loads (HCOD, HNOD, HDOD); point source waste loads (WCOD, WNOD and WDOD); tributary load (TCOD, TNOD, TDOD); and distributed source of CBOD and CNOD. The only variable which is not allowed to be randomized is the travel time, t.

5. Application of FOEA and MCS Techniques to A Nonpoint Source Water Quality Model

The one-dimensional, steady-state water quality model, SNSIM, is capable of simulating CBOD, NBOD and DO concentration, and evaluating the error term associated with each variable, using either FOEA or MCS technique, under the given condition of uncertainty in a multi-reach river system. To demonstrate the application of the methods, it is used for a hypothetical river basin (Braster and Chapra, 1973).

The river basin (Figure 2) consists of five headwaters, three point wasteloads from municipalities and industries, and three tributaries. A withdrawal is treated as a tributary with negative loads in the model. Three upper reaches receive substantial amount of nonpoint source loads of CBOD and NBOD and lower reaches have oxygen demands due to sediments and algal activity. Wasteload No. 1 has a 5000 pounds each per day of CBOD and NBOD.

Thirty-three Monte Carlo runs were conducted with various assumptions of uncertainties in input variables and the discharge from wastewater treatment plant No. 1. Due to lack of actual data, all random input variables were assumed normally distributed with the standard deviations in the range from zero to 40% of the mean values. Each run consisted of 200 simulations.

The 33 MCS runs required 891 sets of two hundred randomized values. The first three statistical moments were evaluated for each set of random values. The means and standard deviations of randomly generated values did not significantly deviate from input values to the randomization process. The average of 891 skewness coefficients was +0.022. Eight hundred and sixteen values were within ± 0.3 and only seven values were outside the range of ± 0.5. Based on this result, the random number generation process performed as desired.

Distribution of DO Values: The minimum DO concentration in the river system occurred between river miles 26 and 28 depending, to some extent, on the waste loads from WWTP no. 1. The range of the minimum concentration was from 3.5 to 5.5 mg/l. The DO concentration at the end of the last reach was raised close to saturation concentration, in the range from 7.5 to 8.5 mg/l for 29 runs. The distributions of DO values at the sag point and at the river end were fitted to normal and log-normal distributions using five goodness-of-fit test (**GOFT**) statistics (Bobba and Singh, 1995).

At the sag point, the normal distribution hypothesis for the DO values was not rejected by any statistics at the 85% confidence level (**CL**) in 25 out of 33 runs, and rejected by all statistics at 97.5% CL in only two runs. The log-normal distribution was not rejected at the 85% Cl in 17 runs and rejected at 97.5% CL in 2 runs.

At the stream end, the normal distribution was not rejected at the 85% CL in 19 out of 29 runs, and rejected at the 97.5% CL again in 2 runs. The log-normal distribution was not rejected at the 85% CL in all 9 runs analyzed.

The MCS performed GOFT tests for log-normal distributions only when the untransformed data are positively skewed and so not contain any negative values. At the DO sag point, either DO or DO deficit data were positively skewed and the GOFT test for log-normal distribution was performed on possitively skewed data. However, at the river end, as DO values approach the saturation value, some DO deficit values were negative due to photosynthetic oxygen production. When DO data were negatively skewed and DO deficit was negative in any simulation, the GOFT test for log-normal distribution was not performed.

Based on this result, it can be concluded that both normal and log-normal distributions describe the distribution of DO values in the stream fairly well. However, because the normal distribution is more widely used and because it out performed the log-normal distribution in a small number of simulations, it would be the distribution of choice in this test.

6. Comparison of FOEA and MCS Results

Two series of computer runs were conducted to compare the output of FOEA and MCS. In one series, the CBOD and NBOD loads from the WWTPS were randomized. These two input variables are linearly related to the DO values in the river. In other series, three reaction rate coefficients, k_c, k_n, and k_a were randomized. These coefficients are nonlinearly related to the DO values.

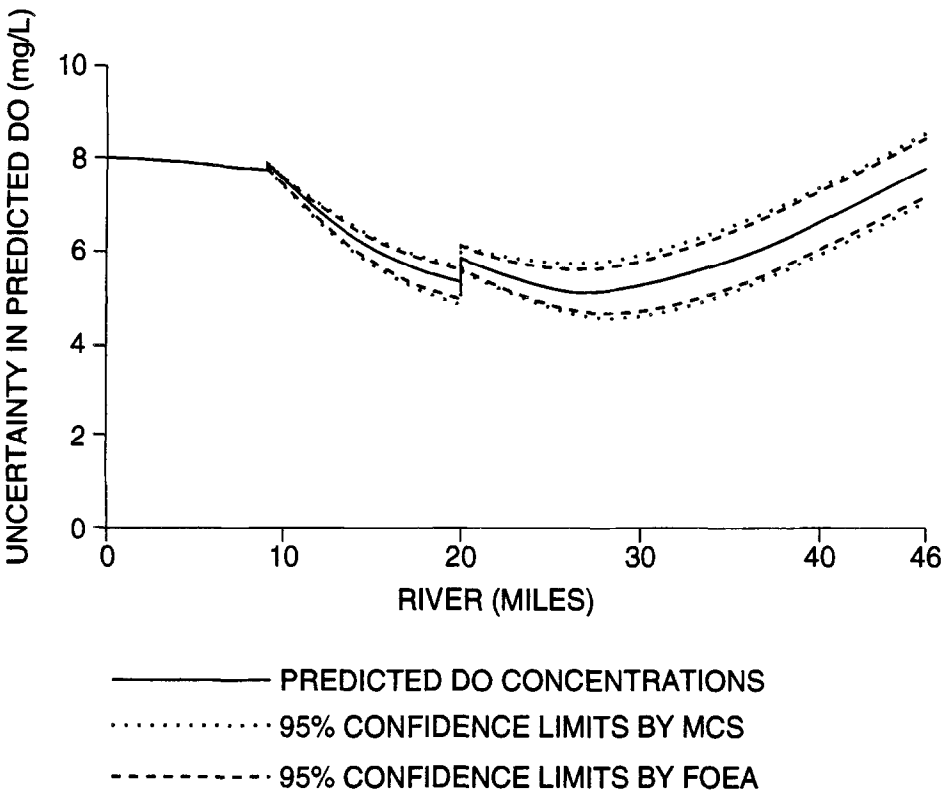

Figure 3. Predicted DO concentration and the 95% confidence limits along the main stream when point source wasteloads are randomized.

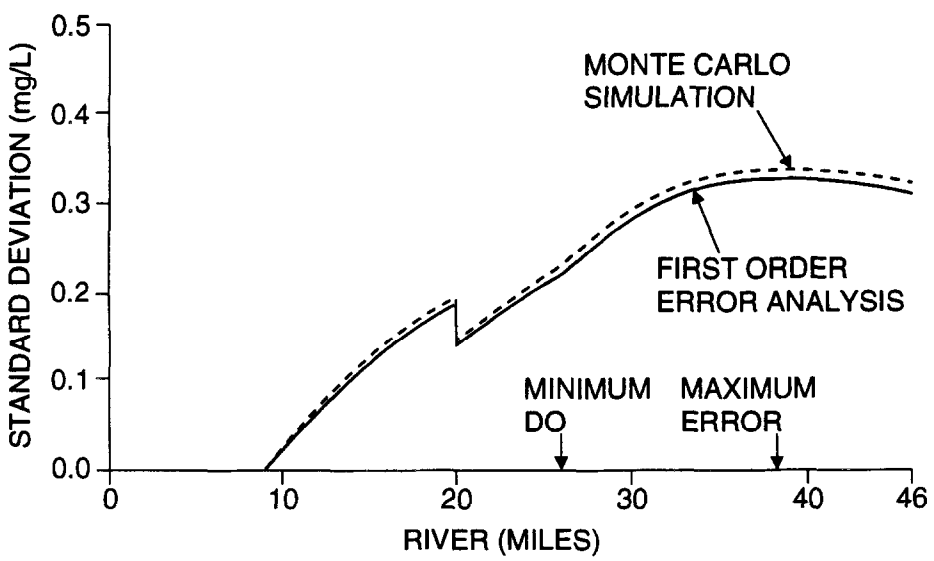

Figure 4. Uncertainty in predicted DO value along the main stream when point source wasteloads are randomized.

(a) Randomization of Linear Input Variables: The predicted DO values and 95% confidence limits generated by FOEA and MCS along the main stream are shown in Figure 3. The standard deviations of the effluent load from WWTP is 20% of the mean value. For this series, WWTP No. 1 was assumed to discharge an average of 2500 pounds/day each of CDOD and NBOD. As expected for the linear input variables, the variations in DO due to the variations in waste loads were in close agreement as predicted by two methods. MCS appears to estimate a slightly larger variation than FOEA, but the maximum difference is less than 3% as shown in Figure 4. The maximum variation occurs beyond the DO sag point.

The GOFT test on the two hundred simulated DO values by MCS at the sag point did not reject the hypothesis of a normal distribution at 85% confidence level. In such a case, the probability of a DO value at the sag point being less than 5 mg/l was evaluated using three methods as follows:

(1) Calculate the standardized value of the minimum required DO concentration (5.0 mg/l in this study) using the mean and the standard deviation of DO values at the sag point predicted by MCS as:

$$Z_5 = \frac{(5.0 - \mathrm{DO})}{S_{\mathrm{DO}}} \tag{9}$$

then read the probability for Z_5 from the standard normal distribution table (parametric method).

(2) Count the number of simulation which predicts DO value at the sag point less than 5.0 mg/l and divide it by the total number of simulations (nonparametric method).

(3) Repeat (1) using the values predicted by FOEA.

Figure 5 depicts the probability of violation (DO < 5 mg/l) vs. the coefficient of variation of the effluent loads. MCS and FOEA produce essentially identical results in the range of variation presented in the Figure 5.

(b) Randomization of Nonlinear Input Variables: In this series of MCS runs, three reaction rate coefficients, k_c,k_n, and k_a were assumed subject to variations. Figure 6 shows the FOEA predicts much wider confidence intervals than MCS, especially near the DO sag point. However, beyond the sag point, as the DO value approaches the saturation value, the errors predicted by two methods become more equal and both generally decrease. This can be more clearly seen in Figure 7 where the standard deviations of the output arrays, using both methods, are shown diverging to a maximum in the sag zone and the converging and decreasing beyond the sag zone.

(c) Figure 8 shows the probability of DO values at the sag point being less than 5 mg/l. The two methods results in a large difference in predicting the probability of DO being less than 5 mg/l, over the entire uncertainty range of input variables, from 0.1 to 0.4 of the coefficient of variations. However, the difference in two methods does tend to decrease as the uncertainty in the reaction rate coefficients

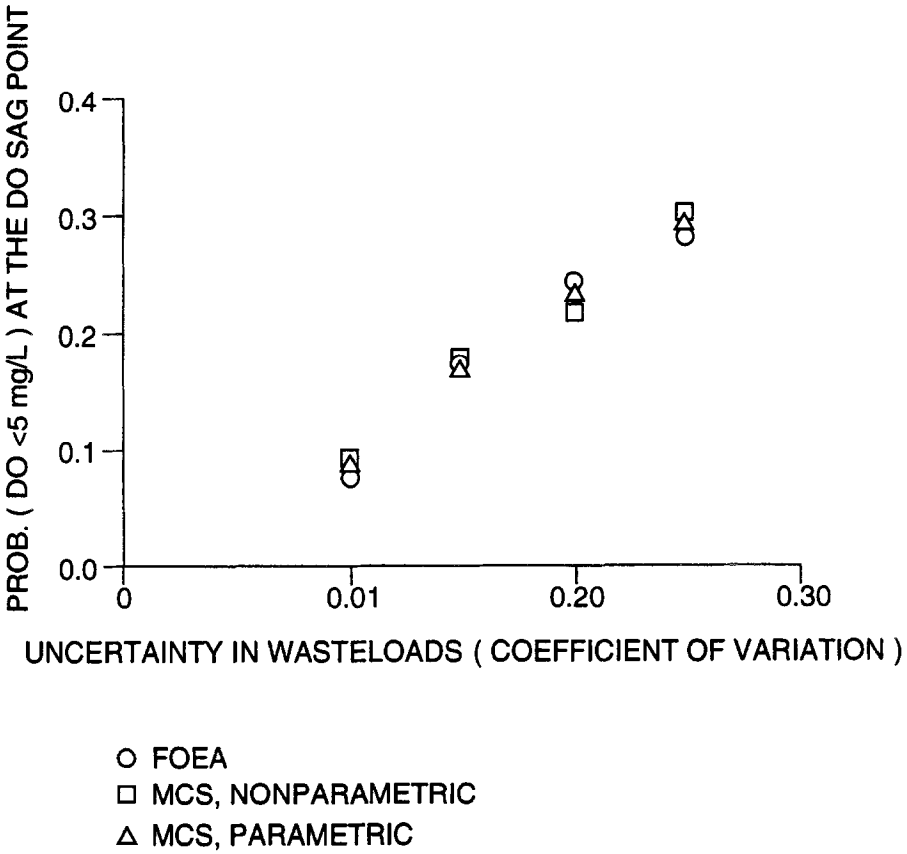

Figure 5. Probability of violation (DO_{sag} < 5 mg/l) when point source wasteloads are randomized.

increases. Since the average DO value at the sag point is 5.2 mg/l (Figure 3), the maximum probability of DO being less than 5 mg/l cannot exceed 0.5 with an estimated normal distribution, regardless of the uncertainty level in input variables.

7. Parameter Estimation

The sensitivity of a model to given input parameter or variable depends on the model structure as well as the probability distribution of the uncertainty in the parameter. In most of the simulation models, a few parameters usually exhibit major influence on the model output. The choice of a value is critical for these parameters, but not for others. Moreover, the degree of influence of a certain parameter may change with temporal and spatial variation. As a result, if an analyst is interested in the model prediction at a certain point, both in time and space, refining the values of one or two parameters which exhibit the most influence at that point may be needed.

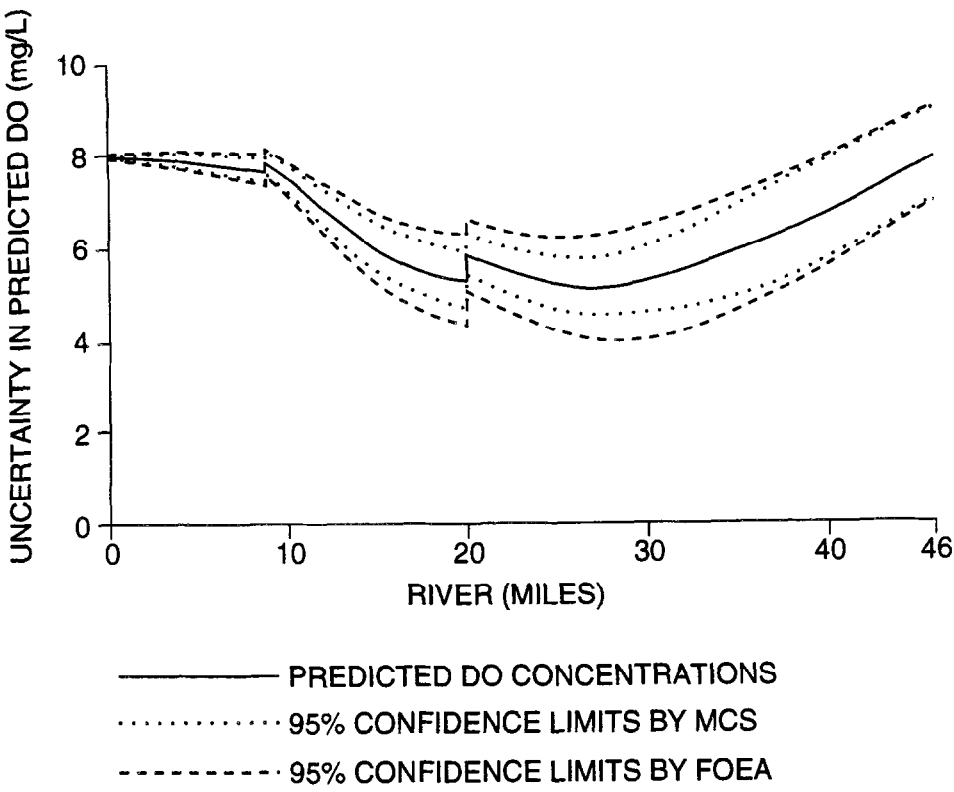

Figure 6. Predicted DO and the 95% confidence limits when k_c, k_n and k_a are randomized.

One advantage in using FOEA is that it provides the components of the total uncertainty in the model output resulting from the uncertainty in each parameter and input variable. The uncertainty components can therefore be combined to give the total uncertainty, or several subtotals of partial uncertainties of similar types.

In the following example, sixteen input variables were exposed to uncertainty; the variables and their uncertainty levels are listed in Table I. These variations might be much smaller than expected, but they still produce significant uncertainties in the predicted DO values.

The uncertainty components in the predicted DO deficiency, in mg/l, contain the combined effect of the uncertainties in the DO deficiency in headwaters, wasteload effluent, and tributary flows, together with the components for CBOD and NBOD. The minimum DO value in the stream is 5.15 mg/l at mile point 26 and the maximum total uncertainty in DO is 0.704 mg/l at the last point in the river. The extent of influence by each input variable on the uncertainty in the predicted DO changes with spatial variations depending on the characteristics of each variable. For example, a DO deficit in the headwaters has its maximum influence at mile point zero.

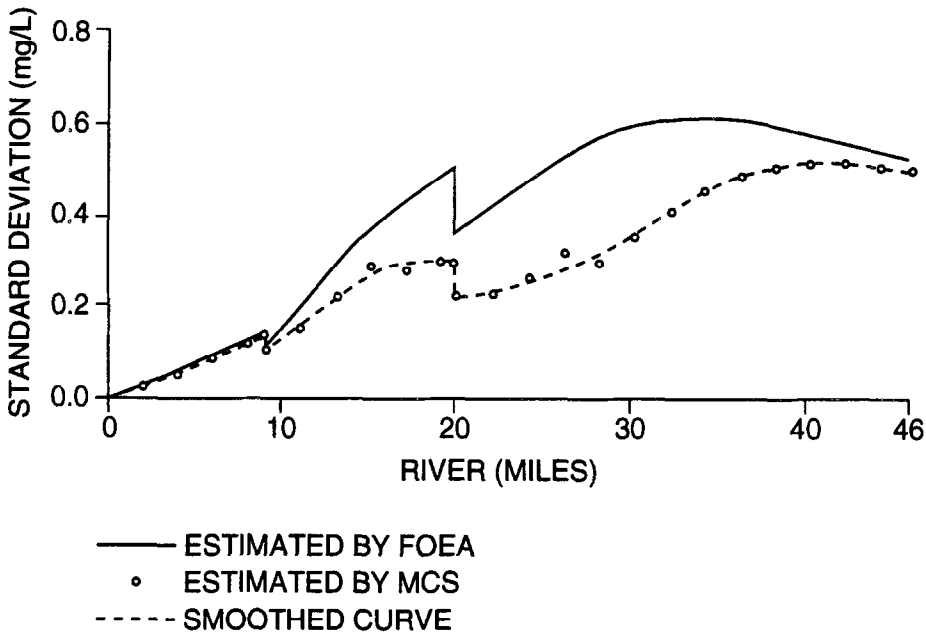

Figure 7. Uncertainty in predicted DO value along the main stream when K_c, K_n and K_a are randomized.

Table I. Uncertainties in input variables for test problem

Variables	Coefficient of variation
CBOD reaction rate coefficient, k_c	0.1
NBOD reaction rate coefficient, k_n	0.1
Atmospheric reaction rate coefficient, k_a	0.1
Headwater CBOD	0.1
Headwater NBOD	0.1
Headwater DO deficit	0.1
Effluent loads CBOD	0.1
Effluent loads NBOD	0.1
Effluent loads DO	0.1
tributary loads CBOD	0.1
Tributary loads NBOD	0.1
Tributary loads DO deficit	0.1
Algal oxygen demand	0.2
Sediment oxygen demand	0.2
Nonpoint source CBOD	0.2
Nonpoint source NBOD	0.2

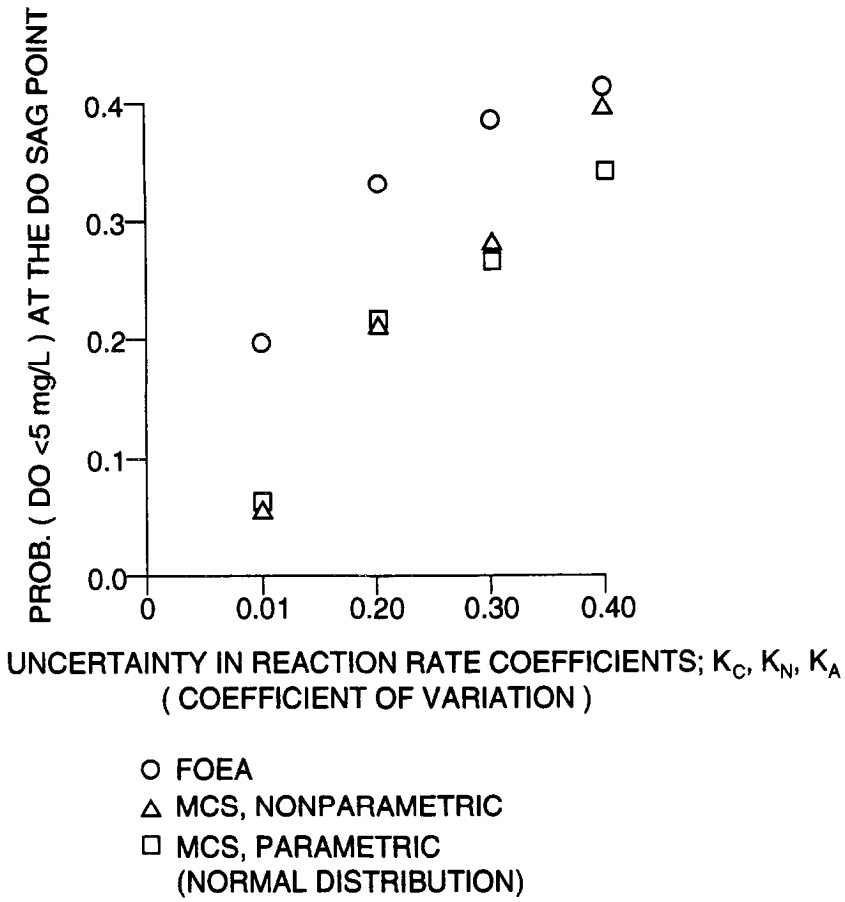

Figure 8. Probability of violation (DO_{sag} < 5 mg/l) when K_c, K_n and K_a are randomized.

Figure 9 illustrates the variation of the uncertainty components in three reaction rate coefficients along the main stream. Note that the square of subtotal uncertainty in DO caused by these three variables in the sum of the squares of uncertainty in the DO caused by each variable, separately.

To present uncertainty components in the predicted DO graphically, the uncertainty components are combined into six subgroups. As shown in Figure 10, at zero mile point, the measurement error of DO (or equivalently DO deficit) in the head water is the only component. However, at the sag point (mile point 26), the uncertainty in CBOD reaction rate coefficient, k_c, is the most important factor. Finally, at the end point, the uncertainty in algal oxygen demand accounts for 71% of the total uncertainty in DO.

If the uncertainty in the algal term were reduced by half, total uncertainty in the DO forecast would decrease by 4% as shown in Figure 11. This example indicates that the method to reduce uncertainty should include increased survey and analysis to better define the more important parameters at the site of interest.

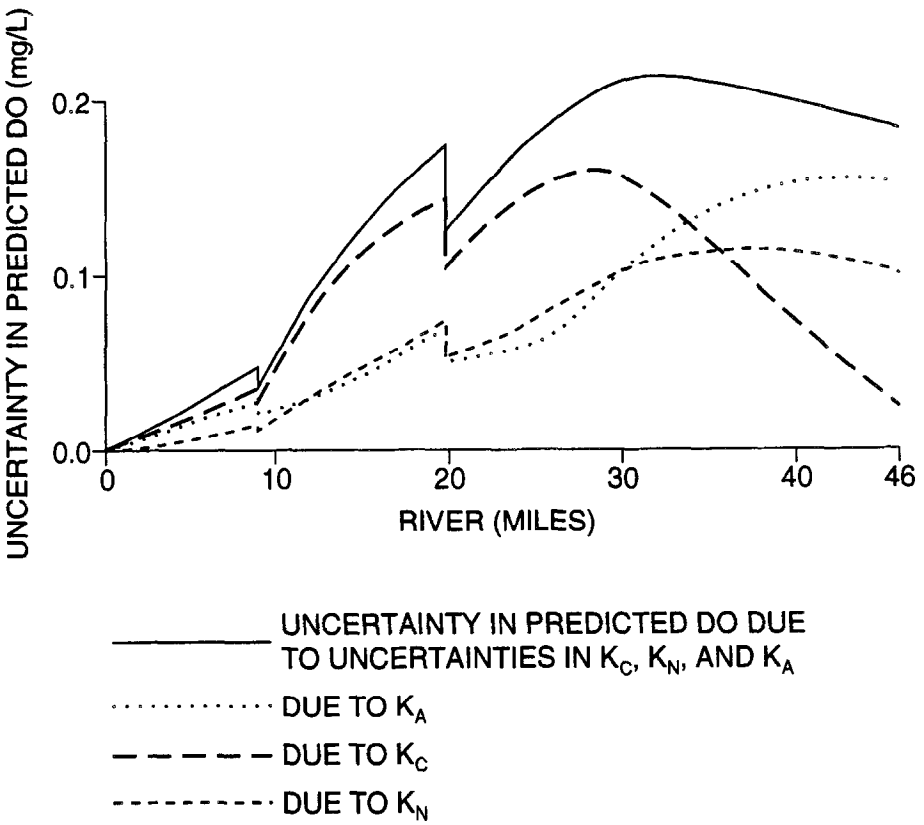

Figure 9. Uncertainty components in predicted DO due to uncertainties in k_c, k_n and k_a.

8. Wasteload Allocation

In the following example, WWPT No. 1 is the largest point-source load discharging 1 million gallons a day containing 5000 lbs/day each of CBOD and NBOD. To improve the downstream water quality, three alternatives are assumed possible for the plant's effluent. For each alternative, the probability of the DO value at the sag point less than the target value 5 mg/l is calculated for increase in the effluent load standard deviation from zero to 25% of the mean value. Figure 12 shows of this situation.

Although alternative B allows the discharge of more total BOD, it would cause less of a probable violation than alternative A at the same level of uncertainty. Alternative C significantly reduced the probability of a violation, but it is noteworthy that 25% variation in alternative C is as vulnerable to violation as only 7% variation in alternative A or B.

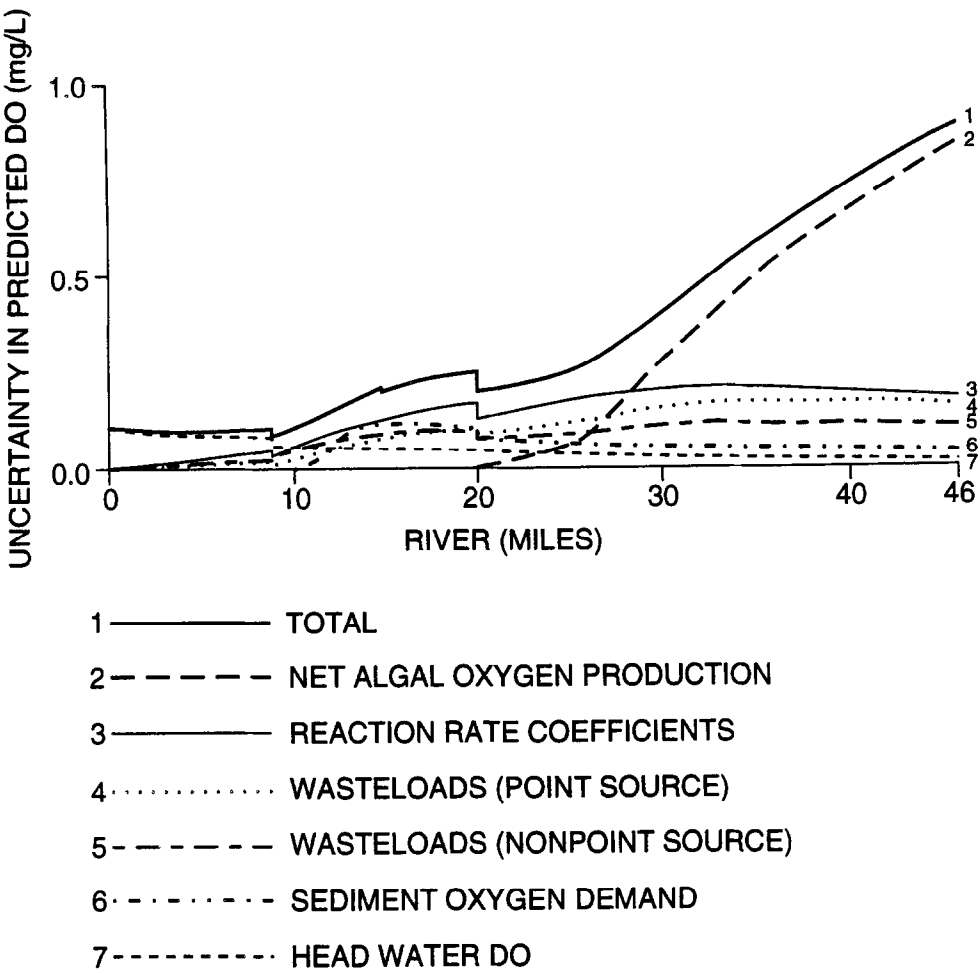

Figure 10. Uncertainty components in predicted DO when 16 inputs variables are randomized.

9. Limitations of Methods

9.1. FUNCTIONAL ANALYSIS

Functional analysis are in general, limited to generation of only the mean and variance of the output variable. Complete description of a distribution requires definition of all higher order moments. Thus, these methods do not entirely describe the distribution concept in the special case where the output is normally distributed. First order error analysis can be extended by truncating the Taylor Series expansion after the second or third order, but carrying higher order terms complicates the mathematics.

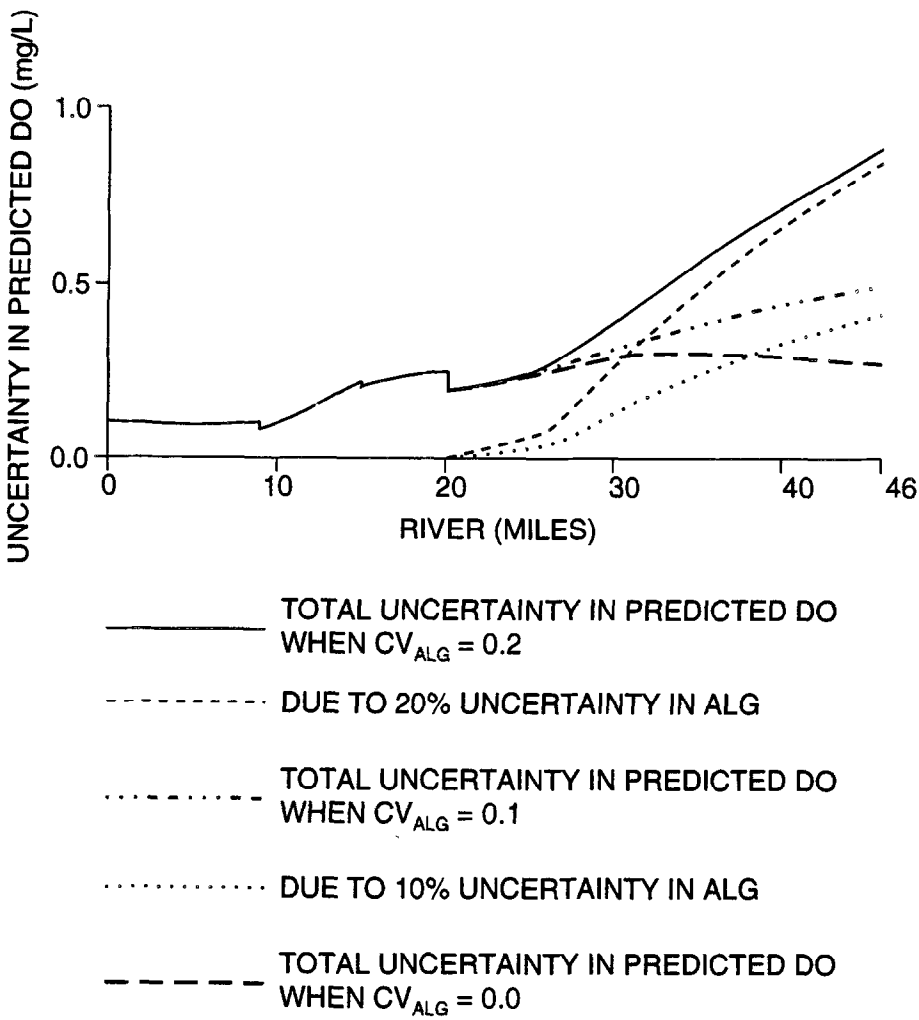

Figure 11. Variation of total uncertainty in predicted DO due to different level of uncertainty in algal oxygen production.

9.2. MONTE CARLO SIMULATION METHOD

While the Monte Carlo method is straightforward, and certain situations shows weakness and limitations. Random number generation, the basis of the method, is not an easy process, and can be troublesome and even unreliable, especially on smaller computers. Random number generators usually produce a set of uniformly distributed values between 0 and 1. If some other distribution for the input variable is required, then the distribution of the random number must be transformed. Detailed discussion of the mathematics of the random variable transformations may be found in Benjamin and Cornell (1970) and Sobol (1974).

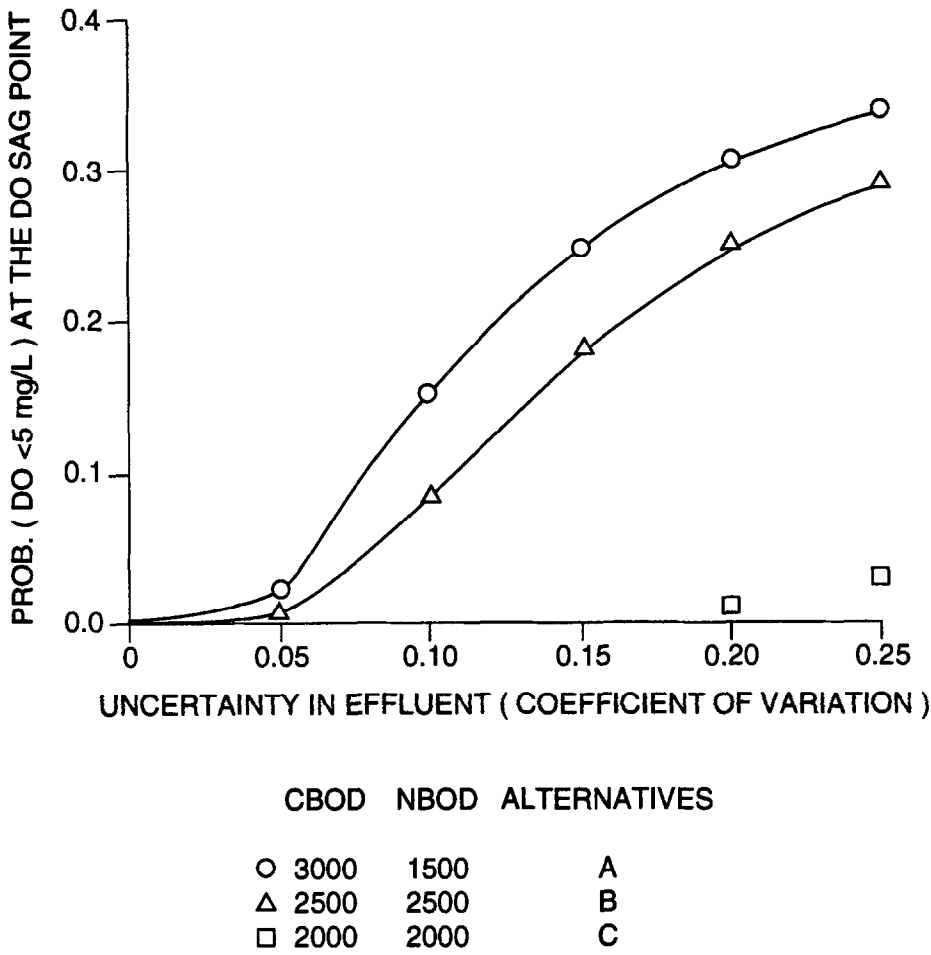

Figure 12. Probability of violation (DO < 5 mg/l) as a function of uncertainty levels in point source loads.

Another central issue in Monte Carlo Simulation methods is determination of the number of times to run the model. Enough runs must be performed to obtain a statistically valid sample. The number required is usually a function of the output distribution, which is not known a priori. In practice, a certain number of trials are executed and the resulting statistics calculated. The process is then repeated with a larger sampling, new statistics calculated and compared to the previous run. If there is no 'significant' difference, then three hundred runs are sufficient. If there is a significant difference, the number of runs are increased and the new statistics compared to the run of six hundred. This process is repeated until no 'significant' difference is found between two successive runs. 'Statistical significance' can be determined by any one of several standard statistical tests for comparison of distributions (Benjamin and Cornell, 1970).

10. Conclusions and Recommendations

In any mathematical modelling, there are inevitable uncertainties in the model structure and parameter input values. The conventional deterministic modelling approach lacks mechanism for estimating the effects of uncertainties on the model response. Sensitivity analysis provides some insight, but is limited. In this paper two error analysis techniques applied in modelling process, first-order error analysis (FOEA) and Monte Carlo simulation (MCS), are discussed in detail and some examples of their use are provided. The following is summary of the conclusions contained in this paper.

(a) The applicability of the first-order error analysis (FOEA) is inherently restricted by three factors: (1) degree of nonlinearity in the model, (2) distribution of errors in each random variable, and (3) magnitude of errors. (b) one particular advantage of the first order error analysis is that it provides the components of total uncertainty in model output caused by each random input variable. When the model is highly nonlinear, first-order error analysis may need to be extended to some higher order-error analysis. (c) Monte Carlo Simulation is conceptually simple, theoretically sound, and flexible. Two important factors to be considered in any application of Monte Carlo Simulation are: (1) probability distribution for each random input variable needs to be selected, and (2) number of simulation requires specification. (d) These two error analysis procedures were incorporated, as options, in a BOD-DO model. This new model is capable of simulating water quality in a complex stream system as well as estimating uncertainties in predicted BOD and DO values which would result from uncertain input conditions. (e) The result of applying these procedures to the model is summarized as follows: (1) If the random input variables are linearly related to BOD and DO in the stream, such as wasteloads from treatment plants, and their distributions are normal, FOEA provides an exact solution for errors associated with predicted BOD, DO values. (2) If a limited number of nonlinear input variables are randomized, MCS provides the option of choice. (3) When a reasonably large number of input variables are randomized, the model output tends to be normally distributed. If moderate departures from normality are present in the random input variables, FOEA provides good estimates of uncertainties in model response. (4) The maximum uncertainty in predicted DO value occurs well beyond the DO sag point for the example selected. (5) The magnitude of influence of uncertainty in each random variable varies with the location along the stream. (6) One or two random variables exhibited a major influence on uncertainty in predicted DO value at any particular point in the stream for the example considered. (7) These error analysis procedures can be used to estimate the probability of violation of minimum DO requirements in stream. (8) The error analysis should be considered in effluent management programs.

Based on the findings in this simulation study, the following recommendations are made: (a) In this study, the modified model was applied to a fictious data set. For further verification, it is desirable to apply this model to data from actual river

systems to further explore the utility of this modelling strategy. (b) This model may be further improved using procedures such as random number generations from 3-parameter gamma distribution, 4-parameter beta distribution, and data-defined and user defined distributions. (c) A general purpose goodness-of-fit test procedure is required to determine the distribution of input variables to available data. It should conclude extreme-value, 3-parameter gamma, and 4-parameter beta distributions.

Current modelling work being performed at NWRI is being directed towards an evaluation of water quality more closely resembling naturally found watershed conditions, and results from these modelling efforts should be shortly forthcoming.

Acknowledgements

We thank Dean Jeffries and G. Padmanabhan, North Dakota State University, Fargo, N.D., U.S.A., for helpful comments on early drafts of this paper. The encouragement and support from John Carey is also appreciated. We would also like to thank the reviewers for their detailed and thoughtful comments.

References

Beck, M. B.: 1987, Water quality modeling: A review of the analysis uncertainty, *Water Resour. Res.* **23**, 1393–1441.

Benjamin, J. R. and Cornell, C. A.: 1970, *Probability, Statistics and Decision for Engineers*, McGraw-Hill, New York.

Black, T. C. and Freyburg, D. L.: 1987, Stochastic modeling of vertically averaged concentration uncertainty in a perfectly stratified aquifer, *Water Resour. Res.* **23**, 997–1004.

Bobba, A. G. and Singh, V. P.: 1995, Application of Monte Carlo analysis of saturated subsurface water contamination modelling. *Proc. Int. Conf. on Hydrology and Water Resources and Water Resources*, Kluwer Academic Publishers, Dordrecht.

Braster, R. E. and Chapra, S. C.: 1973, Documentation for SNIM1/2, a computer program for the steady state water quality simulation of stream network, USEPA, Region II, 26 Federal Plaza, New York.

Burgers, S. J. and Lettenmaier, D. P.: 1975, Probabilistic methods in stream water quality management *Water Resour. Bul.* **11**, 115–130.

Cornell, C. A.: 1972, First-order analysis of model and parameter uncertainty, *Proc. Int. Symp. on Uncertainties in Hydrology and Water Resources Systems*, University of Arizona, Vol. 2., pp. 1245–1272.

Dobbins, W. E.: 1964, BOD and oxygen relationship in streams, *J. Sanitary Eng. Div.* ASCE **90**, 53–57.

Dresnack, R. and Dobbins, W. E.: 1968, Numerical analysis of BOD and DO profiles, *J. Sanitary Eng.* **94**, 789–807.

Fedra, K.: 1979, A stochastic approach to model uncertainty. A lake modelling example. Working paper No. WP-79-63, International Institute for Applied Systems Analysis, Vienna, Austria.

Gardner, R. H., Huff, D. D., O'Neill, R. V., Mankin, J. B., Carney, J., and Jones, J.: 1980, Application of error analysis to a march hydrology model, *Water Resour. Res.* **16**, 659–664.

Gardner, R. H., O'Neill, R. V., Mankin, J. B. and Kumar, D.: 1980, Comparative error analysis of six predator prey models. *Ecology* **61**, 323–332.

Hahn, G. J. and Shapiro, S. S.: 1976, *Statistical Models in Engineering*, Wiley, New York.

Kahn, H.: 1950, Random sampling (Monte Carlo) techniques in neutron attenuation problems–Part I, *Nucleonics* **6**, 27–33.

Karmeshu and Lara-Rosano, F.: 1987, Modelling data uncertainty in growth forecasts. *Appl. Math. Modelling* **11**, 62–68.

Kothandaram, V. and Ewing, B. B.: 1969, A probabilistic analysis of DO-BOD relationships in streams, *J. Water Pollut. Control Fed.* **41**, R73–R90.

van de Kramer, S.: 1983, Monte Carlo Simulation and first order error analysis: Two possible methods to cope with uncertainties in water quality modelling. Working paper No. WP-83-9, International Institute for Applied Systems Analysis, Vienna, Austria.

Malone, R., Bowles, D. S., Windham, M. P., Grennney, W. J.: 1983, Comparison of techniques for assessing of loading uncertainty upon a longterm phosphorous model, *Appl. Math. Modelling* **7**, 11–18.

Melching, C. S.: 1992, An improved first order reliability approach for assessing uncertainties in hydrologic modelling, *J. Hydrology* **132**, 157–177.

Melching, C. S., Yen, B. C., and Wenzel, H. G., Jr.: 1990, A reliability estimation in modelling watershed runoff with uncertainties, *Water Resour. Res.* **25**, 2275–2286.

Metropolis, N and Ulam, S.: 1949, The Monte Carlo Method, *J. Amer. Statis. Associ.* **44**, 335–341.

Meyer, H. A.: 1956, *Symposium on Monte Carlo Methods*, Symp. the University of Florida, 1854, J. Wiley, New York.

Rubinstein, R. Y.: 1981, *Simulation and Monte Carlo Method*, Wiley, New York.

Scavia, D., Canale, R. P., Powers, W. F. and Moody, J. L.: 1981, Variance estimates for a dynamic eutrophication model of Saginaw Bay, Lake Huron, *Water Resour. Res.* **17**, 1115–1124.

Scavia, D., Powers, W. F., Canale, R. P. and Moody, J. L.: 1981, Comparison of first order error analysis and Monte Carlo simulation in time dependent lake eutrophication models, *Water Resour. Res.* **17**, 1051–1059.

Smith, L. and Schwartz, F. W.: 1981, Mass transport. Part 2. Analysis of uncertainty in prediction, *Water Resour. Res.* **17**, 351–369.

Smith, L. and Schwartz, F. W.: 1981, Mass transport, part 3. Role of hydraulic conductivity data in prediction, *Water Resour. Res.* **17**, 1463–1479.

Sobal, I. M.: 1974, *The Monte Carlo Method*, The University of Chicago Press, Chicago.

Spear, R. C.: 1070, The application of Kolmogorov-Renyi statistics to problems of parametric uncertainty in system design, *Int. J. Control* **11**, 771–778.

Thoburn, D.: 1986, A Baysian approach to density estimation, *Biometrika* **73**, 65–75.

Thomann, R. V.: 1972, Systems analysis and water quality management, *Environ. Res. Appl.*

Thomann, R. U.: 1982, Verification of water quality models, *J. Environ. Eng. Div. ASCE* **108**, 923–940.

Tiwari, J. L. and Hobbnie, J. E.: 1976, Random differential equations as models of ecosystems: Monte Carlo Simulation approach, *Math. Biosci.* **28**, 25–44.

Ualm, S. M. and von Neumann, J.: 1945. Random ergodic theorems, *Bull. Ameri. Math. Soc.* **51**, 660–665.

Whitehead, P. and Young, P.: 1979, Water quality in river systems: Monte Carlo Analysis, *Water Resour. Res.* **19**, 1266–1270.